Sociology Wants Mathematics

by

I. Richard Savage

FSU Statistics Report M 247
ONR Technical Report No. 71

January 1973
The Florida State University
Department of Statistics
Tallahassee, Florida


In addition to ONR support this project was begun under a National Science Foundation Senior Postdoctoral Fellowship at the Center for Advanced Study in the Behavioral Sciences, Stanford, California.
Sociology Wants Mathematics

by

I. Richard Savage
Florida State University
Tallahassee, Florida, U.S.A.

ABSTRACT

The current sociological literature contains many examples where the verbal presentation cries out for a formal (mathematical) structure. There are other examples where sociological writings have explicitly used mathematical models which need clarification and analysis. Twelve examples are given of these wants of mathematics.

These examples can serve as an introduction for sociologists and mathematicians of the potential role of mathematics in sociology. Some of the examples suggest major research efforts, such as, the further development of social psychology by more complete formal models and other examples suggest major developmental programs, such as, a decision theoretic approach to population prediction.

The examples do not yield new areas of mathematics coming from sociology. Clearly, many of the published papers could have been clarified and strengthened by modest mathematical efforts in their writing and editing. But the full benefits of mathematical analysis usually would require a substantial effort to bring the available mathematical tools to productive development of the sociological content.
Sociology Wants Mathematics

by

I. Richard Savage
Florida State University

Introduction*

Sociology research, pure and applied, often has mathematical content, I and XII. Examples will be given illustrating two meanings of "want", that is, sociologists who desire to use more mathematics and sociological research which lacks complete success because of insufficient use of mathematics, VII and XI. Illustrating the desire and need for an increase in the use of mathematics in sociology is done to bring about that increase.

"Mathematics" will be used as a description of idealized behavior. Thus mathematics strives for clarity through the consistent use of notation carefully matched to explicit definition. The distinction between logical error and controversy is maintained. Mathematics sometimes will strive for the abstract but mathematics will avoid the vague. Mathematical writing involves formulation of precise problems and attempts to solve those problems. The goals of a mathematical presentation are specified; the reader's need to guess at meanings is held to a minimum. A mathematical document should be simple and elegant; it can be read in

*In the introduction, Roman numerals refer to the examples.
a literal manner. The full power of mathematics is its description of a subject in definitions and axioms from which many properties of the subject are deduced. But the mathematical description of a subject - the first step in the use of mathematics - is difficult and critical to progress.

The illustrations should have the qualities of mathematical writing. As a whole this essay is not mathematical for an orderly formulation and solution of a problem is not attempted. Even so, sufficient material has been examined to convince me that study of such illustrations is worthwhile to help the student, the editor, and the scientist to learn more potent ways of developing and presenting ideas.

It is doubtful that any reader will wish to digest all of the examples. As a collection they can be used as a reading list for a course showing the role of mathematics in standard, current sociology. The collection here might not be adequate, for example, the Journal of Mathematical Sociology is not represented. In picking examples, my interest was directed towards the sociology. My selection process did not allow contributions from those who are primarily mathematicians. In fact, I tended to exclude those contributions which would be commonly regarded as mathematical sociology.

In several of the examples, I have tried to suggest important areas for a joint effort between sociologists and mathematicians. The development of formal models in social psychology is an intellectual challenge
with possible major scientific payoff, IX. Using predictions instead of projection in demography could have major economic and political consequences, VI.

Nearly all of the examples were originally stochastic or I have given them a stochastic formulation, II. No doubt this reflects my interests and training. Still, I think this collection of examples gives strong evidence of the ever present need for the stochastic element in the formulation of sociological theory. In none of the examples have I stressed the need for good experimental design and statistical analysis, III, IV, VIII, X. That neglect was to maintain a reasonable size for this effort.

Although I am most interested in discovering new mathematics, there appears to be little if any new basic mathematics suggested by these examples. Again, this may correspond to my own intellectual inability to spot the new ideas. But I have no evidence that new areas of mathematics have arisen from sociological contexts. On the other hand, I strongly feel that if existing mathematical models are adequate, that it is better to avoid the novel. If nothing else, unnecessary novelty usually brings little errors and unfortunate nuances, XII. Some of the examples have a surprising lack of formal statement, V. At least one example has not yielded substantive results in spite of considerable formal effort, VII.
In the examples I have given a quotation as the center of discussion. Although these quotations may not summarize the article, I think the quotations are characteristic of the article and the quotations are not taken out of context.

The examples are arranged, roughly, in terms of the explicit mathematical content of the original example. Thus, the first examples require much mathematical effort to bring them into a mathematical framework. The last examples were originally presented as mathematical models. The work at hand is clarification and extension of the given model. (The notation and numbering of displays is done independently for each example with some effort to be consistent with the original publications.)


The authors summarize their theoretical models by

A. Interclass and interracial value, attitudinal, behavioral; and experiential differentials on the part of the parental generation conducive to the induction of CNS dysfunction relative to the filial generation (A Sociocultural Variable).

B. The consequent induction of CNS dysfunction (A Biological Variable).

C. The emergence of perceptual-motor problems and/or intellectual deficits symptomatic of the underlying CNS dysfunction (A Psychological Variable).
D. The eventual onset of various kinds of academic "learning disabilities," such as reading problems, related to the perceptual-motor difficulties and/or the intellectual deficits (An Educational Variable).

E. Consequent academic underachievement (An Educational Variable).

F. The determination of reduced socioeconomic status consequent upon academic underachievement - largely by means of occupational determination, in the case of the male and by means of marital selection, in the case of the female (A Sociological Variable).

G. The final induction of an emotional-motivational set, state, or "frame of mind" - consequent upon progressive, continuous, or ongoing socialization in a given region or sector of the social (class) structure - conducive to the development of the values, the attitudes, the behavior, and the experiences associated with factor "A" (A Social=Psychological Variable). (p. 109-110)

Even the most casual reading of the above suggests that the unscrambling of nature and nurture must be very difficult in this context, because of the self-perpetuating of the nurture elements. No attempt is here made to formalize the complete model. Instead, however, a partial formalization of a portion of the model is given.

Let D be the event "damage" and B stand for race-class where {b} are the values of B and large b correspond to the disadvantaged. Then assumptions A. and B. yield:

\[ P(D_t | B_{t-1} = b) \geq P(D_t | B_{t-1} = b') \] when \( b > b' \).

In (AB) the subscripts "t-1" and "t" refer to "time" to be measured in generations. It is sufficient (and appropriate) to think of one family.
Thus $B_{t-1}$ is the race-class of parents and $D_t$ is the event of one of their children being damaged.

From C., D., and E. define an event $U$ which corresponds to lack of personal achievement. Then

$$P(U_t | D_t) > P(U_t).$$

Then F. and G. suggest

$$P(B_t > b | U_t) > P(B_t > b).$$

The authors suggest (p. 110) that this is a "self-perpetual chain", that is,

$$P(D_t | D_{t-1}) > P(D_t).$$

We give an example to show (AB), (CDE) and (FG) do not imply (1).

Assume

$$P(B_0 < b) = P(B_1 < b) = \begin{cases} 
0 & b < 0 \\
0 & 0 \leq b \leq 1 \\
1 & 1 \leq b,
\end{cases}$$

$$D_0 = \{ \frac{1}{3} \leq B_0 \leq \frac{7}{12} \},$$

$$D_1 = \{ \frac{1}{2} \leq B_0 \leq 1 \},$$

and

$$U_0 = \{ \frac{1}{3} \leq B_0 \leq 1 \} = U_1.$$
For this example,

(6) \[ \frac{1}{4} = P(D_0) < P(D_1) = \frac{1}{2} . \]

In this example \( B_0 \) determines everything. We do not think that is realistic and when the example is extended to cover several generations, such a determinism would certainly not be appropriate.

Now let us check the example

(AB) \[ P(D_1 | B_0 = b) = \begin{cases} 1 & \text{if } b \geq \frac{1}{2} \\ 0 & \text{if } b < \frac{1}{2} \end{cases} , \]

(CDE) \[ \frac{2}{3} = P(U_0) < P(U_0 | D_0) = \frac{3}{12} / \frac{3}{12} = 1 , \]

(CDE) \[ \frac{2}{3} = P(U_1) < P(U_1 | D_1) = 1 , \]

(FG) \[ (1-b) = P(B_t > b) < P(B_t > b | U) = \frac{3}{2} (1-\max (b, \frac{1}{3})) . \]

Finally (1) fails as we have

(7) \[ \frac{1}{2} = P(D_1) > P(D_1 | D_0) = \frac{1}{4} \left( \frac{1}{12} \right) = \frac{1}{3} . \]

Apparently, a formal statement of the thoughts behind this paper are possible. To do enough formal analysis to obtain useful results would be costly.

The author summarizes

Proposition 1: The higher the level of the disposing conditions, the greater the likelihood of a collective disturbance.

Proposition 2: Increases in the level of the disposing conditions lead to increases in the amount of administrative control.

Proposition 3: Increases in the amount of administrative control lead to decreases in the level of disposing conditions.

Proposition 4: The amount of administrative control is directly proportional to the efficiency of the control process.

Proposition 5: The total amount of administrative control is limited by the capacity of the control process. (pages 187 and 190)

Even without the author's motivating discussion, one can easily begin to think of the situation as a Markov control process. In that context "administrative control" would be a control or exogenous variable. "Efficiency" refers to the probability of an administrative control procedure being carried out. The author assume the "disposing conditions" will stochastically increase at constant low levels of administrative control.

In preparing notation for the propositions it would be necessary to place a time index on each variable. The propositions can be read as conditional probability statements with the time of the conditioning variable earlier than the conditioned variable.
It would be difficult to develop the detailed formal model along with the necessary operational definitions of the variables to be measured. The qualitative properties of the resulting model might have direct sociological interest. It is doubtful that sufficient data are available to test the model. But no doubt similar models would be appropriate in situations with larger amounts of data.

III


The author observes from his data

Although worker rates decline in the higher parity intervals, there is nonetheless a clear positive relationship between the length of a parity interval and the probability of a woman working in the interval. (page 178)

He then says

From a policy point of view this finding may have considerable significance. As will be shown later in this paper, there is a very strong relationship between whether one works or not in one parity interval and working probabilities in the next higher order interval. (page 178)
The above observations are not surprising. The first observation appears almost necessary. Under what circumstances could one use such observations for policy purposes?

IV


The authors state

Since a new administration takes office in different years, and for different time periods, determining the leadership effect on performance is complicated by each company's unique set of administrative patterns. Hence, we used a simple analytical method to find the proportion of total variance accounted for by each independent variable. For a given dependent variable, say sales, visualize a huge matrix in which each cell contains a specific company's sales figures for a specific year. The sum of the squared differences between each cell and the grand mean, divided by the number of cells, is the total sales variance for the matrix. The influence of any classified variable, say year, may be examined by first determining each category's mean, in this case all companies' mean sales in a given year. Each yearly mean is then subtracted from the matrix grand mean to determine the necessary adjustment for specific sales figures in that year. When the variance of these adjusted scores is determined, the differences between the initial total variance and the adjusted variance may be divided by the initial variance to determine the proportion of total variance eliminated after taking year into account. Similarly, this procedure may be used for other independent variables to determine the percent of total variance accounted for by each. (page 121)
Although some readers might be able to understand and use the verbal description, a standard analytic description through linear models and the analysis of variance would be helpful to many. This analysis might yield more results with greater precision.


The authors basic statement is

In order to check the idea of a "tipping point", we will look at schools with varying proportions of Negroes at the beginning of a particular school year and then see how whites react in those schools. Using this approach we can answer the following question: for all schools with approximately the same percentage of Negroes during one year what was their average increase in proportion in the following year? If we find a point or proportion at which there is an exaggerated increase the following year in the proportion of Negroes, then we will have found a "tipping point", or a point of acceleration in the resegregation process. If we find continuous acceleration of the decrease of whites, we will have supported the social distance notion. (page 129)

The serious reader must first establish some notation: Let $p(t)$ be the proportion of black students and $q(t) = 1 - p(t)$ be the proportion of
white students in a school year $t$. The data are from Baltimore schools (1955-1965) and thus the authors implicitly assume:

$$(1') \quad p(t + 1) - p(t) = -(q(t + 1) - q(t)) > 0.$$ 

Since they speak of "average increase", it would be appropriate to present here a stochastic model. The $p$'s and $q$'s would then be expectations. For this discussion we will assume the situation is deterministic and replace $(1')$ with

$$(2') \quad p(t + 1) - p(t) = f(p(t)) > 0.$$ 

We assume the schools are homogeneous so that $f(\cdot)$ does not depend on other factors, such as, age of school or number of pupils in the school. Note,

$1 \geq p + f(p)$ so that $f(1) = 0$ and $f(p)$ is near $0$ for $p$ near $1$.

A "tipping point" corresponds to a proportion, $p_o$, with an "exaggerated increase", $f(p_o)$. Roughly, $p_o$ is a tipping point if $f(p_o)$ is much larger than values of $f(\cdot)$ at values of $p_o$ slightly less than $p_o$. But "exaggerated increase" is not defined. Will we all agree about the presence of a tipping point? Consider,

Example A.
The discount factor at $p = \frac{3}{4}$ and $f\left(\frac{3}{4}\right) = 2f\left(\frac{1}{4}\right)$ strongly suggests $\frac{3}{4}$ is a tipping point. I think there are no other tipping points. If Example A. is modified so that there is a sloping line up to $(\frac{1}{4}, \frac{3}{4})$, is there a tipping point?

Example A.

\[ f(p) \]

\[ 0 \quad \frac{1}{4} \quad \frac{1}{2} \quad \frac{3}{4} \quad 1 \quad p \]

Is $\frac{1}{2}$ a tipping point? If $p(0) = 0$ then $\frac{1}{2}$ is a tipping point. If $p(0) \neq 0$, $\frac{1}{4}$ nor $\frac{1}{2}$ then $\frac{3}{4}$ is never observed.

Another way of looking at "tipping points" is to consider the "acceleration" which presumably should be measured by

\[ g(p_t) = p(t + 2) - 2p(t + 1) + p(t). \]

Large negative values of $g(\cdot)$ would correspond to our intuitive ideas of a tipping point. The acceleration for whites will be $-g(\cdot)$. 
The social distance notion will be supported by "continuous acceleration" of \( g(\cdot) \). "Continuous accleration" appears to mean \( f(\cdot) \) is continuous and \( g(p_t) = f(p_{t+1}) - f(p_t) \) does not take on large negative values.

The authors have used two key phrases, "exaggerated increase" and "continuous acceleration" without even giving informal definitions. It is doubtful that the title could be answered from data. From the scientific editorial viewpoint, it is surprising that these definitions were not stressed.

VI


Using existing theory is often difficult. Frequently there must be a substantial time period to develop methodology to implement mathematical theory. (There is an analogy to physics and engineering or to
biology and medicine.) Further, there can be misunderstandings, institutional organization, and political pressures which slow down the use of new theory. A striking example of these problems arises in demographic analysis of future populations.

To set the stage, it is first necessary to define population projection and population prediction. To clarify the problem, these definitions will be given in extreme (but not misleading) forms.

A population projection is an extrapolation by formal rules with completely specified assumptions of a future population from a completely specified current population. The projection is deterministic rather than stochastic. Although one should be reasonable and show responsibility in making assumptions, there is no implication that the projection is likely to be realized or that the projection should be used for any phase of decision making.

A population prediction (or forecast) is a statistical estimate of a future population. The estimation procedure utilizes all available knowledge about current data and the likelihood of future events. The estimate may take the form of a point estimate, a confidence interval, or a probability distribution over the class of future populations. Predictions tend to single out the most likely and predictions can be used for making optimal economic decisions.
Projections are relatively easy to compute and it is relatively easy to communicate the method of making projections. Developing and explaining predictions requires the use of much statistical theory. Predictions often are embarrassing since they require a statement of current knowledge. Unlike projections, predictions are to be used by decision makers.

Projections rather than predictions are now used by official demographers. For example, Siegel remarks on the U. S. Bureau of the Census

All the official projections have been conditional projections, and the conditions have been spelled out in detail; none of the series has been offered as forecasts or predictions. They have been designed as deterministic models, i.e., specific or definite values are produced by the model. Although a range is presented for each set of projections, probabilities have not been assigned to the various series in each set. (page 53)

But users of projections will insist on more as Siegel notes and Keyfitz stresses

But the user is typically not a professional demographer and he wants a prediction of what will actually happen in the future. The calculation of future populations, presented as an innocent (indeed tautological) projection by its author, is accepted as a prediction by the reader. The bridge between the two points of view is in the choice of the assumptions. Insofar as these are realistic the projection does indeed forecast what will happen. Hence the stress that will be laid in what follows not only on the nature of the assumption made, but on their presentation in such a way that the user is genuinely in a position to decide on their merits. (page 347)
Muhsam gave an outline of the usefulness of statistical decision theory in making demographic predictions. To date most demographers have not attempted to fill in the full outline. The efforts which have been made tend to concentrate on one aspect of the problem, such as, Schweder applying branching process theory to the development of human populations while not considering errors in the data, lack of stationarity, etc. These efforts have

1. Shown that demographic prediction is difficult.
2. Not convinced most demographers that they are side-stepping their responsibility or that a better (prediction instead of projection) job is feasible and
3. Not developed basic ways for analyzing stochastic aspects or economic consequences of demographic predictions.

A major problem of the demographers is to summarize his knowledge of data in the form of probability statements, that is, he must express his knowledge of the error structures of census reports, mortality tables, etc. He must also summarize knowledge about anticipated changes by probability statements, for example, his probability distributions for future birth rates. These difficult problems must be faced and solved. Careful statements of what is needed and honest effort should work here. To some the problems are so numerous that they give a
response of frustration or ridicule. If an honest effort is made we are more likely to have useful predictions than if the problems are avoided.

In summary, the massive demographic prediction problem wants a solution. Existing theory may be adequate for the task. But bringing the theory to the application requires an effort and conditions which may be difficult to obtain.

VII


In his 1970 abstract Blau (1970) states

The endeavor in this paper is to construct a systematic theory of differentiation in organizations consisting of two basic generalizations and nine propositions derivable from them, which can account for a considerable number of empirical findings. The two basic generalizations are: (1) increasing organizational size generates differentiation along various line at decelerating rates; and (2) differentiation enlarges the administrative component in organization to effect coordination. (page 201)

The abstract is a fair representation of the mathematical formality.

The methods for obtaining the "derivable" propositions are most informal. This article cries out (wants) for a more formal development which was quickly supplied (in part) by Hummon and Meyer with Blau (1971) rejoining.

Hummon establishes the notation

(1) $S$, "size" of organization

(2) $D = f(S)$, amount of "differentiation" in a particular "line".

Typically, "size" is number of employees, a "line" is a classification scheme, such as, skills or levels of supervision, and "differentiation" is the number of classes present in the line. Blau's (1970) first basic generalization is then stated by Hummon as

$\frac{df(S)}{dS} > 0$,

(M1.0a)

$\frac{d^2f(S)}{dS^2} < 0$.

(M1.0b)
With this explicit statement one can ask precise questions:

a. Although (M1.0a) was not explicit in Blau (1970) was it implicit?

b. In (M1.0 a and b) would it make a substantive difference if > and < were replaced by ≥ and ≤? Would this replacement make the mathematical analysis more or less convenient?

c. Since S and D are usually integer valued, is the use of derivatives in (M1.0 a and b) an oversight or an approximation? For what values of S should (M1.0 a and b) be true?

d. The data for the generalizations in Blau (1970) as well as the expository development make clear that the amount of differentiation is a random variable with its distribution function depending on size. Thus, does f(S) represent an expected value, median, etc.?

To finish our discussion we will present the first of the derived propositions in the spirit of Hummon and then suggest what happens when f(S) is thought of as a random variable.

We need

\[(2)\]

Let C be the average size of component, that is

\[C = S/f(S), \, S > 0.\]

Then the first derived proposition is

\[(3)\]

If \(f(0) = 0,\)

(M1.0a)' \(f' > 0, \, 0 \leq S\)

(M1.0b)' \(f'' < 0, \, 0 \leq S\)
right hand derivatives at \( S = 0 \) - then

\( (M1.2)' \quad C' > 0, \quad 0 < S. \)

**Proof.** Begin with

\( (A1.2)' \quad C' = (f - Sf')/f^2 \)

so that \( (M1.2)' \) is equivalent to

\( (4) \quad f - Sf' > 0, \quad 0 < S. \)

Notice

\( (f - Sf')' \)

\( (5) \quad = f' - f' - Sf'' = -Sf'' > 0, \quad 0 < S \)

where the last inequality follows from \( (M1.0b)' \). But the last inequality with \( (f-Sf') \big|_{S=0} = 0 \) implies \( (4) \).

Hummon implicitly uses \( f(0) = 0 \) in the proof of \( (A1.2) \).

Now if \( f(S) \) is thought of as a random variable and \( S \) takes on the values \( 0, 1, \ldots \) we can reformulate the problem. For convenience use

\( (6) \quad E(S) = Ef(S). \)

Then assume

\( (7) \quad f(0) = E(0) = 0, \)

\( (8) \quad f(1) = E(1) = 1, \)

\( (M1.0a)'' \quad E(S) - E(S-1) > 0, \quad S \geq 1, \)

\( (M1.0b)'' \quad E(S) - 2E(S-1) + E(S-2) < 0, \quad S \geq 2. \)
Now define

\[ C(S) = E(S/f(S)) = SE(f(S))^{-1} \]

for \( S = 1, 2, \ldots \).

Then (M1.2) becomes

(M1.2)' \[ C(S) - C(S-1) > 0 \text{ for } S = 2, \ldots \]

Without strong assumptions, one usually finds

(9) \[ E(X^{-1}) \neq (E(X))^{-1}. \]

Thus we can not write \( C(S) = S(E(S))^{-1} \). In fact, one can construct counter-examples to show that (M1.2)' is not a consequence of the assumption which precedes it.

Thus the Blau (1970) work has been put by Hummon into a mathematical form where at least some of his propositions can be derived by standard mathematics. But the deterministic formulation of Hummon is not satisfying and to date no one has given an adequate stochastic formulation of Blau (1970). Mayhew, et al. present a stochastic model which they conjecture satisfies (7), (8), (M1.0a)' and (M1.0b)' but they did not investigate (M1.2)' or the other derivations of Blau (1970).

Begin with a collection of $N$ people. Each of them names his best (first) friend. Say $n$ people are named as best friends and the case of naming yourself as best friend is excluded. Thus we assume $N \geq 2$ and deduce $N \geq n \geq 2$. A reciprocating pair of individuals $(a,b)$ occurs when $a$ names $b$ and $b$ names $a$ as best friends. Let $t = \text{number of reciprocating pairs}$ and the primary concern of the authors is the random variable $T = 2t = \text{number of individuals in reciprocating pairs}$. The authors propose

For the moment let us consider only those persons who are named as a first friend and try to construct a probabilistic model of choice reciprocation. The simplest model would be a sequential choice process in which the person would reciprocate this choice as a first friend with a probability $\phi$. (page 430)

Although the model is elaborated for the discussion of 2nd choices, etc., no further basic statements appear about the model. The authors appear to conclude (pp. 433-434) that $T$ has a binomial distribution with parameters $\phi$ and $n$. (From the text it is not clear if they really mean $t$ or $T$, $n$ or $N$, or exact or approximate distributions.)

Two questions arise: How can one interpret this model and is it mathematically consistent? If the first question is resolved then will $t$ have approximately a binomial distribution?
A first task is to give a precise definition of $\phi$. We try

$\phi = P(a \text{ chooses } b \mid b \text{ chooses } a, \text{ someone chooses } b, \text{ and } n \text{ were chosen from } N)$, where $2 \leq n \leq N$.

The value of $\phi$ can depend on $n$ and $N$ but we are assuming it does not depend on $a$ and $b$ provided $a \neq b$. Notice $\phi = 1$ if $n = 2$ and $\phi < 1$ if $n$ is odd.

Some thought shows there are numerous ways of achieving these conditions on $\phi$, for example,

a. The $N$ people form a circle and each individual chooses the person on his right as his best friend. Then $P(n=N) = 1$ and $\phi = 0$ if $N \geq 3$.

Here $T$ and $t$ have degenerate binomial distributions.

b. Each of the $N$ independently choose (with equal probabilities) one of the others. Then thinking intuitively, $\phi = (n-1)^{-1}$. Also, let $X_{ab}$ be the event that $a$ chooses $b$ and in the following $a$ and $b$ will correspond to the chosen persons. Then

$$E(T|n) = E(\sum_{a \neq b} X_{ab} X_{ba} |n)$$

$$= n(n-1) E(X_{12} X_{21} |n)$$

$$= \frac{n(n-1)}{(n-1)^2} = \frac{n}{n-1}.$$  

(This is much too small for the data of the article.)
c. The $N$ individuals sit in a circle and each independently chooses either (with equal probability) his right or left neighbor as best friend. When $N = 3$, we get, typically,

$$a \ C \ b \ C \ c \ C \ a$$

where $aCb$ means $a$ chooses $b$. So $n = 3$ and $\phi = 0$. Or we could get

$$a \ C \ b \ C \ c \ C \ b.$$ 

So $n = 2$ and $\phi = 1$. If $N = 4$, we can obtain $n = 3$, typically,

$$a \ C \ b \ C \ c \ C \ d \ C \ c$$

and $\phi = \frac{1}{2}$.

We leave it to the industrious reader to find out if these models yield a binomial distribution.

In summary, the postulation of the existence of $\phi$ does not lead to a vacuum. In fact there are many possible models. It is not clear, however, that binomial distribution will yield good approximations for the distribution of $t$ except in degenerate cases. Because $T$ must be even, its distribution will not be approximately binomial except when $P(T = n^\#)$ is near one where $n^\#$ is an even integer.

Finally, the authors may have had an entirely different concept of $\phi$ in their minds. The reader is never sure when he is forced into substantial second guessing.


The assumptions of Moore have an appearance of formality.

Assumption 1: Given S and O as status differentiated actors in ambiguous, task-oriented interactions, S will generalize from differentiated states of the general status characteristic to correspondingly differentiated self-other performance expectations regarding the task at hand.

Assumption 2: At any stage of the interaction process, S associates a disagreement between himself and 0 with differences in evaluations of alternatives by himself and 0.

Assumption 3: At any stage of the process, S will tend to positively evaluate that act made by the actor for whom he holds relatively high performance expectations and negatively evaluate that act made by the actor for whom he holds relatively low performance expectations.

Assumption 4: At any stage of the process, if S positively evaluates one alternative and negatively evaluates the second, then he will select the first and reject the second.
Assumption 5a: Whatever his responses on the first n-1 steps of the interaction process, a high status S will on the nth step of the process positively evaluate his own act with probability \( p \). 
Assumption 5b: Whatever his responses on the first n-1 steps of the interaction process, a low status S will on the nth step of the process positively evaluate his own act with probability \( q \). (pages 147, 148 and 149)

The mathematical readers will want to make these assumptions in a more precise form. Moore suggests probability theory as the appropriate framework. In fact, Moore modifies Assumption 5 to allow "learning" or "non-stationarity".

My experience says that translating such assumptions to precise mathematical statements is possible although difficult. The translation might involve resolutions of ambiguities which make substantive differences, that is, the mathematician must work with the social psychologist. When done with his translation effort, the mathematician may be disappointed because the mathematical model is not rich in consequences. His disappointment when verbalized to the sociologist might result in enrichment of the model, further analysis of current data, and the design and performance of new experiments.

The Moore paper is typical of many papers in experimental social psychology. As other examples we mention Burke, which is slightly less formal than Moore, and Webster, which is slightly more formal than Moore. Savage and Webster carries out the translation program for Webster.
Cohen, Zelditch and Berger (1972) represents a similar but larger challenge than Webster (1969). Some of the writing of Homans have been formalized with varying degrees of success.


Begin with

Define the size \((S)\) of an organization to be the number of individuals (employees, personnel) it contains. External structural differentiation \((\Delta)\) is defined as the number of occupations in the organization. Then, internal structural differentiation \((\delta)\) becomes \(\delta = \Delta/S\). Clearly, \(1 < S < \infty\), and in organizations where one individual can fill no more than one occupational role, \(1 \leq \Delta \leq S\), and \(1/S \leq \delta \leq 1\). (p. 816)

The authors think of \(\Delta\) having a probability distribution and

A minimally restrictive assumption about this distribution, in the absence of specified forces, would be that each possible outcome is equally likely for any given \(S\); therefore, the conceptualized distribution of \(\Delta\) values is rectangular (or uniform). (p. 819)
The meaning of "minimally restrictive" and why it leads to

\[ P(\Delta = k|S) = \begin{cases} S^{-1} & k=1, \ldots, S \\ 0, & \text{otherwise} \end{cases} \]

is not clear. It does not seem particularly plausible:

Note (1') implies

\[ E(\Delta|S) = \frac{(S + 1)}{2} \]

and

\[ V(\Delta|S) = \frac{(S^2 - 1)}{12}. \]

Thus \( E(\Delta|S) \) increases quickly, although it is compatible with Blau (1970). A correct statistical analysis of data from (1') would account for (3').

Without much more effort one could avoid the selection of a specific distribution (1') and consider a family of distributions, such as,

\[ P(\Delta=k|S) = \begin{cases} Ck^{A-1}(S + 1-k)^{B-1} & k=1, \ldots, S \\ 0, & \text{otherwise} \end{cases} \]

and \( C \) is selected so that \( \sum_{k=1}^{S} P(\Delta=k|S) = 1 \). The family (4') contains (1'): let \( A = B = 1 \). The family (4') is analogous to the Beta family which is rich in forms and easy to analyze. The family (4') is relatively simple and thus in the spirit of this paper. If (4') or something of comparable complexity is not adequate, then the next step would be to allow a complexity such as allowing \( A \) and \( B \) to be functions of \( S \).
"Minimally restrictive" is not a well defined concept and (1') might be a poor representation of the thought. One is reminded of Bayesian statistical inference and the inadequate representation of "total ignorance" by uniform distributions.


The author states

At the individual level the rate of change in the probability of getting married would then be:

\[ \frac{dp_i}{dt} = qP_t \tag{1} \]

where \( p_i \) is the probability that individual \( i \) in the cohort will marry in the small time interval \( dt \); \( P_t \) is the proportion of the cohort already married at time \( t \); and \( q \) is the parameter of conversion into the marriage state.

However, it is hard to obtain observations about probability changes at the individual level. Therefore let us translate the individual probabilities into rates of change at the aggregate level.

Let us assume

\[ \frac{dp_i}{dt} = \frac{dp}{dt} \tag{2} \]
for all \( i \) and \( j \). That is, we do not assume the probability of getting married is the same for all cohort members, but that the rate of change in their respective individual probabilities is the same. To get the rate of change in the number married, we add the individual rates of change for those not yet married:

\[
\frac{dm}{dt} = n \frac{dp_i}{dt} = (n-m) \frac{dp_i}{dt} = (n-m)qP_t
\quad i=m+1
\]

where \( m \) is the number married, and \( n \) is the number in the cohort. If we now divide (3) through by \( n \), we obtain

\[
\frac{dm}{ndt} = \frac{d \left( \frac{m}{n} \right)}{dt} = \frac{dP_t}{dt}
\]

\[
= \frac{m-n}{n} qP_t = q(1-P_t)P_t.
\]

(page 174)

Apparently, one should interpret \( dp_i/dt \) in the following manner:

If individual \( i \) is not married at \( t \) then his probability of marriage in the interval \((t, t+dt)\) is \((dt)(dp_i/dt) + o(dt)\) where the remainder is uniform in \( t \). Since \( p_i \) is not used we need not consider its interpretation. In saying \( P_t \) is a proportion, it should be made clear that it is the expected proportion or probability that a randomly selected individual in the cohort has been married at time \( t \), that is, \( P_t \) is not a random variable.

Assumption (2) appears to be unnecessary since it is implied by (1).

In (3) \( m \) is treated both as a discrete variable and as a continuous variable. It seems best to let \( m = m(t) \) be the expected number of people
married by time t. We then have

\[(a) \quad m(t + dt) - m(t) = \sum_{i=1}^{n} \left( P(i \text{ not married at } t) \right) \]

\[= \sum \left( 1 - P_i \right) \frac{dp_i}{dt} dt + o(dt). \]

Then dividing by dt and taking the limit, we obtain

\[(b) \quad \frac{dm(t)}{dt} = \sum \left( 1 - P_t \right) \frac{dp_i}{dt} \]

which yields (4) when we use (1) and divide by n.

The author then says "Next we consider the specific nature of [q=] f(t)". And he selects

\[(6) \quad f(t) = Ab^t \]

which yields an explicit form for P_t. However, we could proceed generally and obtain

\[(9') \quad -\log \frac{1 - P_t}{P_o} = \int_0^t f(s) ds \equiv F(t) - F(o) \]

or

\[(10') \quad P_t = \left( 1 + \frac{(1 - P_o)}{P_o} \right) e^{-\left( F(o) - F(t) \right)} \]

In this more general (than (6)) framework, one can make statements which might have demographic significance. For example, \( P_t \to 1 \) if and only if \( F(t) \to \infty \).

Comment. The author's formulation of the model and derivation of (4) appears obscure but easily clarified. His specialization at (6) may have cut him off from general theory.


Some difficulties are just slips, for example,

> Since the long-run individual matrices $P^*(m)$ are constant. (McFarland, page 470)

From page 469, it is clear the author would like to replace "are constant" by "have identical rows".

Some difficulties are unwise from the applied viewpoint, for example,

There exists some small positive number, $\epsilon$, such that for each person $m$ and for each pair of states $x$ and $y$, $P_{xy}(m) \geq \epsilon$. (page 469)

If we make epsilon sufficiently small, there is no means of reliably inferring within the duration of any one mobility study (or the lifetime of a given man) whether one of a man's probabilities is actually zero, rather than epsilon or greater. (McFarland, page 469)
In many social science applications there will be biological and institutional reasons forcing $P_{ij}(m) = 0$ for some $i, j, m$ combinations, for example, men do not become pregnant, and boys of 16 do not become U. S. senators at 17.

Some difficulties may be suggestive of need for change in problem formulation, such as

let $P(m)$ denote the matrix of transition probabilities for person $m$. We want to combine these transition matrices for the various persons in such a manner as to yield the expected value of the one-step transition matrix for the entire population, i.e., a matrix $Q_{1}$ whose $i$-$j$ element is the expected proportion of the people in category $i$ at time 0 who are in category $j$ at time 1. To do this, we define $N_{0}(m)$ as a diagonal matrix with an entry of unity in the diagonal position corresponding to person $m$'s initial state and zeros elsewhere; and let $N_{0} = \Sigma_{m} N_{0}(m)$ be the diagonal matrix whose $j$-$j$ element is the number of people initially in category $j$. Then $N_{0}^{-1}$, the inverse of $N_{0}$, is a diagonal matrix whose $j$-$j$ element is the reciprocal of the corresponding element of the matrix $N_{0}$. Using this notation, the expected one-step population transition matrix is

$$Q_{1} = N_{0}^{-1} \Sigma_{m} N_{0}(m) P(m). \quad (2)$$

Similarly, the expected $k$-step population transition matrix is

$$Q_{k} = N_{0}^{-1} \Sigma_{m} N_{0}(m) [P(m)]^{k}. \quad (3)$$

(McFarland, page 469)
Some worries are:

1. The existence of $N_0^{-1}$ implies that at time 0, there was at least one person in each category.
   a. Hence the present model is not, as might be anticipated, a generalization of the standard theory for one person.
   b. In working with cohorts in education or occupation studies there will be some unoccupied categories at time 0.

2. In the theory of Markov chains the transition matrix is defined without reference to the initial distribution or the initial category of a person. But here, although $Q_1$ is called the "expected one-step population transition matrix" it depends on the initial states of the persons through the values of $N_0(m)$.

3. The meaning of "expected" in phrases such as "expected one-step population transition matrix" is obscure since there is no indication of which probability distribution is used in forming the expectation.

It is hard to guess what an author really meant or wishes that he said. But let us examine some of the possibilities.

The element in the $(x,y)$ - position of $Q_1$ is

\[ Q_1(x,y) = \sum_{m} P_{x,y}(m) / N_x \]

where $N_x$ is the $x^{th}$ coordinate of $N_0$, that is, $N_x$ is the number of persons initially in the $x^{th}$ category and $P^x$ is summed over those persons initially in the $x^{th}$ category. For a fixed $x$ we can interpret the row vector.
(B) $Q_1 x = (Q_1(x,1), Q_1(x,2), \ldots, Q_1(x,y), \ldots)$ in the following manner:

Consider a two-stage process: First select with equal probabilities one of the $N_x$ persons initially in the $x^{th}$ category. Then let the selected person, say $m_0$, make a move according to his transition matrix $P(m_0)$. For this process, $Q_1 x$ is the absolute or marginal distribution after one step of the location of an individual initially in state $x$. These are not conditional probabilities, the usual components of a transition matrix. Notice $Q_1 x'$ does not give us information about what would have happened to individuals who in fact were initially in $x$ if those people had been in $x'$. Thus $Q_1$ and $Q_k$ can be interpreted probabilistically, the offered interpretation is to consider their rows as marginal probabilities. The matrix structure of $Q_k$ is not of interest.

In the McFarland model the initial states of persons and their particular transition matrix, $P(m)$ are mixed together. Here we propose a new model. We hope it includes the kinds of phenomena of interest to McFarland and at the same time its interpretation will be clearer.

(C) Let $P(m)$, $m = 1, 2, \ldots$, represent Markov matrices. Different types of people will have different Markov matrices but several people can have the same matrix.

(D) Let $T(m)$, $1, 2, \ldots$, be the probability that a person has $P(m)$ as his Markov matrix. Thus $T(m) \geq 0$ and $\sum T(m) = 1$. The substantive origin of
T(m) could be from the frequency of occurrence of types in nature or these probabilities could be fixed by experimental conditions.

(E) Let \( p(x|m) \) be the initial probabilities for people with \( P(m) \) as their Markov matrix.

For this model

(F) \( P^k(m) \) is the \( k \)-step transition matrix for the people of the \( m^{th} \) type. Assume \( \lim_{k \to \infty} P^k(m) \) exists and call it \( P^\infty(m) \).

(G) \( (p(\cdot|m))P^k(m) \) is the (row) vector of \( k \) step marginal probabilities for people of type \( m \). Here

\[
(p(\cdot|m)) = (p(1|m), \ldots, p(x|m), \ldots).
\]

Note limit \( \lim_{k \to \infty} (p(\cdot|m))P^k(m) = P^\infty(m) \)

where \( P^\infty(m) \) is a row of \( P^\infty(m) \), all the rows being identical.

(H) \( \Sigma p(x|m)T(m) \) is the probability of starting in category \( x \).

(I) \( \Sigma T(m)P^k(m) \) is a matrix whose \((x,y)\) element is the conditional probability that a person selected with probabilities \( T(\cdot|\cdot) \) from the population will be in \( y^{th} \) category after \( k \) steps given that he started in category \( x \). This matrix will have a limit with identical rows.

(J) \( \Sigma_m p(x|m)T(m)P_{xy}(m) \) is the probability of being initially in category \( x \) and after one step in category \( y \). These are not conditional probabilities.

Note even if \( P_{xy}(m) \) did not depend on \( x \) - consider \( P^\infty_{xy}(m) \) - the matrix defined at (J) will not have identical rows.
\\( (k) \quad \sum_m \frac{p(x|m)T(m)P_{xy}(m)}{\sum_m p(x|m)T(m)} \)

will be denoted by \( Q_{xy} \). Note this is the conditional probability of being in category \( y \) given the initial category was \( x \). The choice of notation, \( Q_{xy} \), was appropriate. Consider the case where each \( P(m) \) corresponds to a different person. The \( T(m) = 1/N \). \( \sum_m = \Sigma_{0} \) in McFarland's notation. Then \( p(x|m) \) is 1 if the \( m \)th person was initially in category \( x \) and otherwise \( p(x|m) = 0 \). Thus the formula \((J)\) becomes

\[
\frac{\sum_{m} P_{xy}(m)}{\sum_{m} 1}
\]

which is another form of \((A)\).

In applications we would wish to infer from observations reasonable values for \( T(\cdot) \), \( p(\cdot|\cdot) \) and \( P \). How to proceed will depend on the available information and the kinds of data.

Spilerman (1972, page 289, footnote 17) notes that McFarland should have considered conditions for the existence of \( N_{0}^{-1} \). He also notes that the terminology of McFarland is not standard (page 278, footnote 1). But Spilerman finds some intuitive reason to continue the non standard terminology.

It is my hope that the above discussion will help make some precision possible in future research in this area. Even more, I think the detail given here will be suggestive for future work. Thus the above model permits
each person to be a different type but the idea of types of persons should in some areas greatly reduce the number of parameters in the population. In the detailed model there are some nice probability problems even after we know the parameters. For example, if after $k$ (or many) steps a person is in category $y$, what is the probability that he is of type $m$.

In closing this note we mention that Stinchcombe (1969) contained several mathematical difficulties. Starbuck (1970) saw that and furthermore made constructive modifications of Stinchcombe.

One side benefit of good writing is that errors can be found, removed, and constructively replaced. Appropriate use of mathematics is part of good writing.
The current sociological literature contains many examples where the verbal presentation cries out for a formal (mathematical) structure. There are other examples where sociological writings have explicitly used mathematical models which need clarification and analysis. Twelve examples are given of these wants of mathematics.

These examples can serve as an introduction for sociologists and mathematicians of the potential role of mathematics in sociology. Some of the examples suggest major research efforts, such as, the further development of social psychology by more complete formal models and other examples suggest major developmental programs, such as, a decision theoretic approach to population prediction.
13. Abstract - continued

The examples do not yield new areas of mathematics coming from sociology. Clearly, many of the published papers could have been clarified and strengthened by modest mathematical efforts in their writing and editing. But the full benefits of mathematical analysis usually would require a substantial effort to bring the available mathematical tools to productive development of the sociological content.

<table>
<thead>
<tr>
<th>Key Words</th>
<th>Link A</th>
<th>Link B</th>
<th>Link C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Role</td>
<td>Role</td>
<td>Role</td>
</tr>
<tr>
<td></td>
<td>Wt.</td>
<td>Wt.</td>
<td>Wt.</td>
</tr>
<tr>
<td>Models</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probability applications</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Markov chains</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sociology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demography</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social psychology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control theory</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>